Final Technical Report

USGS Award Number: G16AP00021
Title: Digitization of Harvard-Adam Dziewoński Analog Seismograms from 1933 to 1953 for Improved Seismicity Constraints in the Northeastern United States

Miaki Ishii
Department of Earth & Planetary Sciences
Harvard University
20 Oxford Street
Cambridge, MA 02138
(617)384-8066
(617)384-8249 (fax)
ishii@eps.harvard.edu

Term: January 1, 2016 to December 31, 2016

Abstract

The goals of this project can be divided into three. The first is the preservation efforts for the analog seismograms, the second is the continued development of the digitization software, and the third component is the use of the software to digitize images of analog seismograms. The first goal has been completed after nearly 6 years of effort by a volunteer. We have made significant progress in the software development, in allowing the user to analyze small portion of the seismogram rather than requiring analysis of the entire image. Even though this sounds minor, it resulted in significant reorganization of how we envision the digitization procedure and reduction in processing time. Finally, continuous efforts are being made to test the software and digitize seismograms with focus on 1938 and 1939.

Report

The main aims of the project was to preserve seismograms from the Harvard-Adam Dziewoński Observatory recorded between 1933 and 1953, and to continue our development of a software that allows robust and highly accurate conversion of the analog images to digital time series. With this grant, we have made significant progress on both aims as described below.

Cleaning and Scanning of HRV Seismograms

This labour-intensive effort has been undertaken by our dedicated volunteer, Hiromi Ishii, and has been completed in 2016. In the end, we were able to save about 10,200 analog seismograms which have been cleaned and scanned. Because the back side of the seismograms contained information about the time, instrument, etc., we have scanned both the front side (where the seismic traces are) and the back side. The front side has been scanned in color with 1200 dpi resolution while the back side has been scanned in gray with 400 dpi resolution. Due to the high resolution, the front images are typically 1.5 GB in size, and in total, the images take up about 41 TB. All the scanned images are available on the seismology web page through the yearly links at
http://seismology.harvard.edu/HRV/scanned_images.html
Over the year, we received a few inquiries about these images, and the most common issue was

the file size. They were too large to be downloaded without very broadband internet connection. To alleviate the problem, we also converted almost all the images to the JPEG format which are about 50 MB each, and provided links through the same web pages. The JPEG files add about 470 GB of data.

After painstakingly checking the scanned images against the original to make sure that we have high-quality scans of every single seismogram, we began the negotiation with Harvard archives for curation of the collection. Working with an archivist, the transfer has been completed, and the analog seismograms between 1933 and 1953 are now under the care of Harvard archives. They are stored in a space that is temperature and humidity controlled, and any researcher interested in accessing the data can do so through a request to the archives section of the Harvard Library.
http://library.harvard.edu/university-archives

Development of Digitization Software

Significant part of the grant was used to partially support a postdoctoral fellow, Petros Bogiatzis, with the efforts to continue development of DigitSeis, a MATLAB based software for processing the scanned image into digital time series. The basic version of the software was completed with the previous USGS grant, G14AP00016, and was provided openly to the community through publication (Bogiatzis & Ishii, 2016), the seismology web site
http://seismology.harvard.edu/research/DigitSeis.html
and at the GitHub repository
https://github.com/PetrosBogiatzis/DigitSeis
With grant G16AP00021, we focused on increasing the usability and flexibility of the software during the digitization stage. The improvements are:

- improved object assignments to traces
- digitization of traces that do not include time marks
- option for the user to select trace(s) to be digitized
- ability to remove very small objects that take up too much memory and crashes the program
- ability to filter image to remove graininess of the image (improves performance by removing small objects and making traces better defined)
- addition of the standard-deviation measure during digitization. This results in two lines being generated for each trace, one for the trace and another for a measure of the width of the trace.
- ability to correct for trace curvature for pen-generated seismograms
- option to correct a portion of trace after digitization

During these improvements, minor bugs have been fixed and improvements have been incorporated. The last item, an ability to correct a small portion of the digitized trace, turned out to be extremely useful. Prior to this addition, the user needed to clean the image very carefully and spend significant amount of time classifying the objects, because once the traces are digitized, there was no way to correct them. If there was anything wrong, the whole image had to be reprocessed. Now, the user spends less time on image quality and classification since small parts of the image can be corrected after digitization. This resulted in substantial decrease in the time user spends on a single seismogram and also reduced the level of frustration.

The next obvious step was to improve the timing component of the software, and that has been completed (with additional improvements) with the new USGS grant G17AP00007. We are currently preparing to release DigitSeis version 1.1 which has all the improvements. The new manual and software will be made available at the seismology web site once they are ready (we hope we will complete the manual in the next month or so).

Digitization

One of the key input for DigitSeis improvements was feedback from the users. The volunteer who has been working on cleaning and scanning of seismogram has been trained to use some functionalities of the software. In addition, with the support from this grant, combined with Harvard University Faculty Aide Program that provides partial funding for employing undergraduate students, an undergraduate student, Thomas Lee, has been hired to digitize seismograms. They have been using the DigitSeis software and providing useful feedback that are reflected in the improvements listed above. Because we encountered similar challenges with timing algorithm of the original DigitSeis as mentioned, we have been focusing on processing seismograms up to the digitization step (the timing step comes after digitization). Between the two users, we have about 80 seismograms from 1938 and 1939 that have been digitized (without timing information). Because there is no timing information, we have not converted these seismograms to SAC files to be made available online, but with the improved timing step, we hope to make these available within the next year.

Bibliography

1. Bogiatzis, P., & Ishii, M., 2016. DigitSeis: A new digitization software for converting analog seismograms. *Bull. Seismol. Soc. Am.* **87**(3), 726–736, doi:101785/0220150246.